

# Predicting Drugs Adverse Side-Effects Using a Recommender-System

Diogo Pinto<sup>2(✉)</sup>, Pedro Costa<sup>2</sup>, Rui Camacho<sup>1,2</sup>, and Vítor Santos Costa<sup>1,3</sup>

<sup>1</sup> INESC TEC, Porto, Portugal

<sup>2</sup> DEI and Faculty of Engineering, University of Porto, Porto, Portugal  
diogojapinto@gmail.com

<sup>3</sup> DCC, Faculty of Sciences, University of Porto, Porto, Portugal

**Abstract.** Adverse Drug Events (ADEs) are a major health problem, and developing accurate prediction methods may have a significant impact in public health. Ideally, we would like to have predictive methods, that could pinpoint possible ADRs during the drug development process. Unfortunately, most relevant information on possible ADRs is only available after the drug is commercially available. As a first step, we propose using prior information on existing interactions through recommendation systems algorithms. We have evaluated our proposal using data from the ADRcS database with promising results.

**Keywords:** Adverse drug effects · Adverse drug reactions · Singular value decomposition · Recommender-systems · Pharmacovigilance

## 1 Introduction

Adverse Drug Events (ADEs)<sup>1</sup> are events that indicate a relationship between the treatment and a negative outcome. It is estimated that, in the United States alone, ADEs account for up to 28 % of all emergency department visits [8], and 5 % of hospital deaths [3]. As a consequence, between 30 and 150 billion dollars are spent annually in hospitals treating those adverse events [4]. There is thus, not only a moral obligation on pursuing safer medicines, but also strong economic impact.

Randomized Controlled Trials (RCTs) are the main tool used to ensure drug quality. They are conducted in standardized conditions, nonetheless, authors have noticed under-representation of women and elderly patients in those trials [6]. Alongside RCTs being conducted regardless of the specific features of the drug or the patient, they often use small samples and with very little statistical significance. Due to these limitations, only ADEs that are common and that develop over short periods of time can be detected with high-confidence.

In this work, we aim at taking advantage of the ability of Machine Learning to process large amounts of data in order to find hidden connections. Our method is as follows. First, we collect data that would be publicly available before the

---

<sup>1</sup> Also referred to as Adverse Drug Reactions (ADR).

drug enters the market. Second, we feed that information to a recommendation system. The output is the set of side-effects with higher estimated probability. We experimented our method, using two methods: Singular Value Decomposition (SVD) and Restricted Boltzmann Machines (RBM), and then combining them as an ensemble classifier.

The main contribution of the present work is the adaptation and evaluation of recommender systems to the problem of predicting ADEs. As the empirical evaluation shows, the technology is scalable and flexible, and enables ADE prediction at any stage of the drug’s development (with special focus on the pre-marketing stage).

## 2 Related Work

Most ADR research is done on the post-marketing stage, where not only there is more information available, but also when large amounts of money were already invested and the cost of discovering a new ADE is considerably higher. Such research has relied on a variety of data sources. One major source has been electronic health records (EHRs), even though they pose challenges of their own [7].

Our approach is inspired on the excellent performance of recommender systems in sparse domains [5]. A significant boost to research in recommender systems was due to the NetFlix challenge. The winning entry of the competition was an ensemble of several algorithms, including various Singular Value models blended with RBM [2]. Our work applies and adapts these methods to the challenging task of ADE prediction.

## 3 Methods and Algorithms

### 3.1 Singular Value Decomposition

**Formalization.** The drug-ADE relationship is represented as a matrix  $M \in R^{m \times n}$ , where  $m$  is the number of drugs and  $n$  the number of ADEs. Whenever a drug  $d$  is known to cause ADE  $a$ ,  $M_{da} = 1$ . This representation causes  $M$  to be sparse.

Matrix factorization allows not only the mapping of drugs and ADEs in factor-spaces but also the reduction of the matrix dimensionality. Consider that each drug is associated with a vector  $p_i$  and each ADE with a vector  $q_i$  such that:

$$M = PQ^T \quad (1)$$

The Singular Value Decomposition (SVD) is a factorization of a real (in our case) or complex matrix. Let’s consider the factorization of a real valued matrix. Formally, the singular value decomposition of an  $m \times n$  matrix  $M$  is a factorization of the form  $M = U\Sigma V^T$ , where  $U$  is an  $m \times r$  orthonormal matrix,  $\Sigma$  is an  $r \times r$  diagonal matrix with positive, non-zero, singular values in

decreasing order, and  $V^\top$  is an  $r \times n$  orthonormal matrix, where  $r$  is the rank of matrix  $M$ . Then, it is possible to obtain matrices  $P$  and  $Q$  from Eq. 1 by:

$$P = U\sqrt{\Sigma} \quad (2)$$

$$Q = \sqrt{\Sigma}V^\top \quad (3)$$

We are interested in matrix  $Q$  of size  $r \times n$ , whose entries represent the “meta” relations between  $r$  pseudo-drugs and the  $n$  ADEs.

**Dimensionality Reduction.** The model generated by the method described above might suffer from over-fitting, since it would fit the noise present in  $M$ . One solution to generalize the model and reduce the effects of the noise is to find a matrix  $\hat{M}$  which is the best rank  $k$  approximation of  $M$ , with  $k < r$ .

The problem to be solved is, then, to find the optimal value for  $k$ . The energy of the factorization of a matrix is defined by Rajaraman and Ullman [9] as the sum of the squares of all its singular values. The new reduced matrix is obtained by discarding a certain amount of that energy.  $k$  is the value that minimizes  $\epsilon$ :

$$\left| \sum_{i=1}^k \Sigma_{ii}^2 - \alpha \sum_{i=1}^r \Sigma_{ii}^2 \right| = \epsilon, \quad (4)$$

where  $\alpha \in [0, 1]$  is the amount of energy we wish to keep. Tests showed that the optimal value of alpha for this problem is 0.9, as greater values reduce the Recall and smaller values reduce the Precision.

**Gradient Descent.** After the dimensionality reduction step, it is possible to optimize  $P$  and  $Q$  by using gradient descent.

$$\min_{P, Q} \sum_{training} (m_{xi} - p_i q_x^\top)^2 + \lambda \left[ \sum_x \|p_x\|^2 + \sum_x \|q_i\|^2 \right] \quad (5)$$

The real goal is to find  $P$  and  $Q$  based on known drug-ADE relations so that we predict well the unseen values. This enables us to approximate missing drug-ADE relations as zeros.

### 3.2 Restricted Boltzmann Machines

**Formalization.** Restricted Boltzmann Machines (RBMs) [10] can be used to perform a binary factor analysis. An RBM is a stochastic neural network consisting on a layer of visible units, a layer of hidden units and a bias unit. The visible units represent, in this context, the drug’s ADEs that we know. The hidden units are the latent factors that we want the model to learn. The visible units and the hidden units form a bipartite graph.

It is possible to reduce the dimensionality of a feature vector, in the case that the hidden layer has fewer units than the visible layer. By providing a drug  $d$  with size  $1 \times n$  it is possible to obtain a vector  $f$  with its latent factors of size  $1 \times l$  with  $l$  being the number of hidden units. On the other hand, vector  $l$  can also be used to obtain a vector  $\hat{d}$ .  $\hat{d}_i$  represents the probability of drug  $d$  causing ADE  $i$ .

### 3.3 Ensembles

SVD and RBM are able to predict the probability of a drug causing a set of ADEs. More formally, given a drug  $d$  of size  $1 \times n$ , each method predicts a drug  $\hat{d}$  of the same size:

$$\hat{d} = [\hat{p}_1 \ \hat{p}_2 \ \cdots \ \hat{p}_n] \quad (6)$$

where  $\hat{p}_i$  represents the probability of drug  $d$  causing ADE  $i$ .

One can look at the problem of making an element-wise combination of  $\hat{d}_{SVD}$  and  $\hat{d}_{RBM}$  as a classification problem where, given two probabilities the model classifies the final probability as positive (*i.e. causing the ADE*) or negative (*i.e. not causing the ADE*); or as a regression problem where, given two probabilities the model computes a new probability.

We have used a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel for the classification problem. On the other hand, to solve the regression problem, a Support Vector Regression (SVR) algorithm was used, also with a RBF kernel. The SVR model was trained the same way as the SVM model, nonetheless, a threshold was required in order to be able to classify a drug as causing ADE  $i$  or not.

The Receiver Operating Characteristic (ROC) curve was computed and the threshold is obtained by using the Youden index [12], which maximizes  $Sensitivity + Specificity - 1$ . Graphically, the index is represented by the maximum height above the chance line.

## 4 Empirical Evaluation

### 4.1 Data

In these experiments we use the ADReCS<sup>2</sup> data-base as ground truth. This drug-ADE database is maintained by researchers at Xiamen University, and includes adverse drugs' reactions ontologies, that enable the standardization and hierarchization of ADE terms. The drug-ADE information of ADReCS was mainly sourced from the drug labels in the DailyMed, maintained by the U.S. National Library of Medicine (NLM) [1].

### 4.2 Methodology

The samples were randomly split into two different sets: 70% into a training set and the remaining into a test set. In the case of the SVD and the regression ensemble, the elements of the resulting prediction vary between 0 and 1. A threshold is needed to distinguish between a positive and a negative example. To do that, the ROC curve is computed by using a validation set, and the Youden index [12] is used as the threshold, as described in Subsect. 3.3. The validation set is computed differently for each method and, therefore, it is explained in the corresponding Subsection. To test the model, the testing set is used. For each

<sup>2</sup> <http://bioinf.xmu.edu.cn/ADReCS>.

element of the testing set, 30 % of the ADEs are randomly removed and used as input to the model. The Precision, Recall and, whenever possible, the ROC area are computed.

**SVD Experiments.** To build the model,  $k$ -fold cross validation was applied to the training set, with  $k = 10$ . At each iteration, 9 folds are chosen as matrix  $M$  and matrix  $Q$  is computed, as described in Subsect. 3.1, leaving the remaining fold as the validation set.

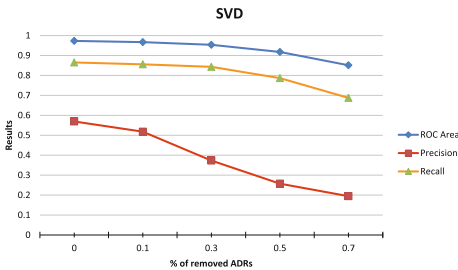
After obtaining the model, the method was tested using the testing set. The computed metrics are presented on Table 1.

**Table 1.** Results of the SVD by removing 30 % of each drug’s known ADEs present in the test set

|                    | ROC area | Precision | Recall |
|--------------------|----------|-----------|--------|
| Average            | 0.954    | 0.373     | 0.843  |
| Standard deviation | 0.054    | 0.120     | 0.136  |
| Minimum            | 0.410    | 0.000     | 0.000  |
| Maximum            | 1.000    | 1.000     | 1.000  |

The system performs well on the majority of the elements from the test set, as can be seen by the large ROC area and small standard deviation. With this approach, on average, the system is able to find 84 % of the ADEs of each drug. Nonetheless, about 37 % of the elements classified as positive are, indeed, positive, as concluded from the Precision.

A sensitivity analysis was performed, by varying the number of removed ADEs from the test set, in order to evaluate the system’s performance under different conditions. As show in Fig. 1, the system’s performance deteriorates as the level of information is reduced.



**Fig. 1.** Sensitivity Analysis of the different metrics by varying the number of removed ADEs from the test set

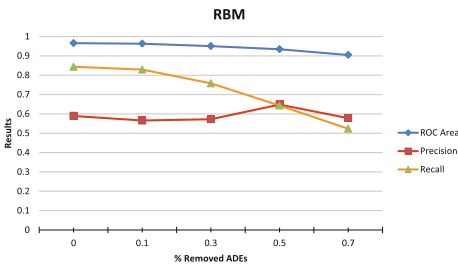
**Table 2.** Results of the RBM by removing 30 % of each drug’s known ADEs present in the test set

|                    | ROC area | Precision | Recall |
|--------------------|----------|-----------|--------|
| Average            | 0.950    | 0.572     | 0.758  |
| Standard deviation | 0.051    | 0.276     | 0.196  |
| Minimum            | 0.417    | 0.000     | 0.000  |
| Maximum            | 1.000    | 1.000     | 1.000  |

**RBM Experiments.** The RBM model was built using 1000 hidden nodes, and was trained for 300 epochs. The results are presented on Table 2. Also, a sensitivity analysis was performed, the same way as for the SVD, and is presented on Fig. 2.

It is possible to conclude that the RBM provides better Precision but with lower Recall than the SVD. Also, this method deals better with the absence of information than the SVD.

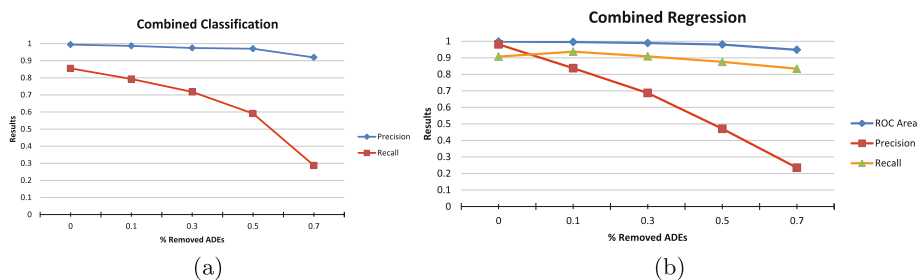
**Ensemble.** In order to combine the two methods, 10% of the training set is used as validation set. The results of combining SVD and RBM using the classification approach and the regression approach are compared on Table 3. Again, a sensitivity analysis was performed on the two approaches and presented on Fig. 3.



**Fig. 2.** Sensitivity Analysis of the different metrics by varying the number of removed ADEs from the test set

**Table 3.** Results of combining SVD and RBM, using SVM and SVR, by removing 30 % of each drug’s known ADEs present in the test set

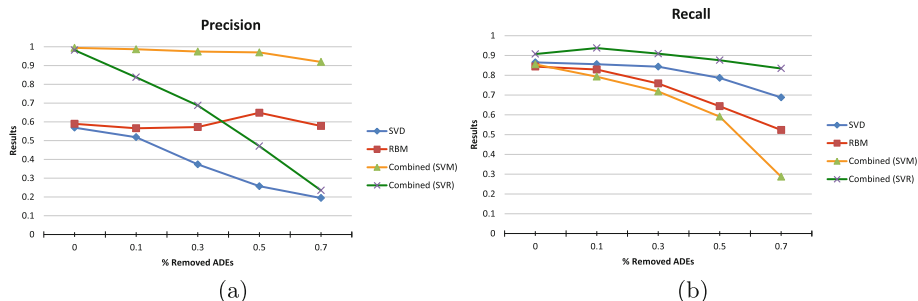
|         | Precision(SVM) | Recall(SVM) | ROC(SVR) | Precision(SVR) | Recall(SVR) |
|---------|----------------|-------------|----------|----------------|-------------|
| Average | 0.974          | 0.718       | 0.990    | 0.687          | 0.909       |
| SD      | 0.112          | 0.198       | 0.036    | 0.128          | 0.140       |
| Minimum | 0.000          | 0.000       | 0.509    | 0.000          | 0.000       |
| Maximum | 1.000          | 1.000       | 1.000    | 1.000          | 1.000       |



**Fig. 3.** Performance of all methods by varying the number of missing ADEs. (a) Precision. (b) Recall.

### 4.3 Discussion

By analyzing Fig. 4 it is possible to conclude that the combined approach is able to maximize one of the two metrics, but not both. The classification one is able to achieve high precision but has low recall, on the other hand, the regression approach achieves high recall but low precision.



**Fig. 4.** Results of using the classification and regression approaches to combine the SVD and RBM models. (a) Classification. (b) Regression.

## 5 Conclusions and Future Work

The method presented here serves as a basis for further expansion. It is capable of taking other data to strengthen its results *e.g.*, molecular descriptors, molecular substructures, literature statistical analysis or even patients information.

Another connection particularly interesting is the comparison of the results with drug→side-effect reports that can be mined from a database such as the FDA Adverse Event Reporting System (FAERS), based on the approach of Rong Xu and QuanQiu Wang [11]. On the other hand, the comparison against other methods, such as different variations of SVM and Random Forests, could give more insight and even boost the precision and recall of the ensemble method.

In sum, there is still much work to be done based on this method, and most important, as a reminder to the research community of the importance on focusing on pre-marketing prediction (and consequent prevention) strategies for ADEs.

**Acknowledgements.** The authors gratefully acknowledge the financial support of Fundação para a Ciência e Tecnologia (FCT), through the research project “ADE - Adverse Drug Effects Detection” (PTDC/EIA-EIA/121686/2010), as well as the Master in Informatics and Computing Engineering (MIEIC) at FEUP.

## References

1. Cai, M.-C., Xu, Q., Pan, Y.-J., Pan, W., Ji, N., Li, Y.-B., Jin, H.-J., Liu, K., Ji, Z.-L.: ADReCS: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms. *Nucleic Acids Res.* **43**, 907–913 (2015)
2. Gower, S.: Netflix Prize and SVD, pp. 1–10 (2014)
3. Juntti-Patinen, L., Neuvonen, P.: Drug-related deaths in a university central hospital. *Eur. J. Clin. Pharmacol.* **58**(7), 479–482 (2002)
4. Lazarou, J., Pomeranz, B.H., Corey, P.N.: Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *Jama* **279**(15), 1200–1205 (1998)
5. Leskovec, J., Rajaraman, A., Ullman, J.D.: Mining of Massive Datasets. Cambridge University Press, UK (2014)
6. Martin, K., Bégaud, B., Latry, P., Miremont-Salamé, G., Fourrier, A., Moore, N.: Differences between clinical trials and postmarketing use. *Br. J. Clin. Pharmacol.* **57**(1), 86–92 (2004)
7. Page, D., Costa, V.S., Natarajan, S., Barnard, A., Peissig, P., Caldwell, M.: Identifying adverse drug events by relational learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 2012, p. 790. NIH Public Access (2012)
8. Patel, P., Zed, P.J.: Drug-related visits to the emergency department: How big is the problem? *Pharmacother. J. Hum. Pharmacol. Drug Ther.* **22**(7), 915–923 (2002)
9. Rajaraman, A., Ullman, J.D.: Mining of Massive Datasets. Cambridge University Press, Cambridge (2011)
10. Smolensky, P.: Information processing in dynamical systems: Foundations of harmony theory (1986)
11. Xu, R., Wang, Q.: Large-scale combining signals from both biomedical literature and the fda adverse event reporting system (faers) to improve post-marketing drug safety signal detection. *BMC Bioinform.* **15**(1), 17 (2014)
12. Youden, W.J.: Index for rating diagnostic tests. *Cancer* **3**(1), 32–35 (1950)